

#4

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of

Yasunaga MIYAZAWA



Group Art Unit: 2121

Application No.: 09/955,164

Filed: September 19, 2001

Docket No.: 110567

For: METHOD FOR CALCULATING HMM OUTPUT PROBABILITY
AND SPEECH RECOGNITION APPARATUS

CLAIM FOR PRIORITY

Director of the U.S. Patent and Trademark Office
Washington, D.C. 20231

Sir:

The benefit of the filing date of the following prior foreign application filed in the following foreign country is hereby requested for the above-identified patent application and the priority provided in 35 U.S.C. §119 is hereby claimed:

Japanese Patent Application No. 2000-285330, filed September 20, 2000.

In support of this claim, a certified copy of said original foreign application:

 X is filed herewith.

 was filed on in Parent Application No. filed .

 will be filed at a later date.

It is requested that the file of this application be marked to indicate that the requirements of 35 U.S.C. §119 have been fulfilled and that the Patent and Trademark Office kindly acknowledge receipt of this document.

Respectfully submitted,

James A. Oliff
Registration No. 27,075

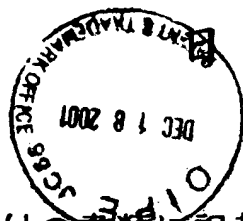
Eric D. Morehouse
Registration No. 38,565

JAO:EDM/gam

Date: December 18, 2001

OLIFF & BERRIDGE, PLC
P.O. Box 19928
Alexandria, Virginia 22320
Telephone: (703) 836-6400

DEPOSIT ACCOUNT USE AUTHORIZATION Please grant any extension necessary for entry; Charge any fee due to our Deposit Account No. 15-0461
--



本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2000年 9月20日

出 願 番 号

Application Number:

特願2000-285330

出 願 人

Applicant(s):

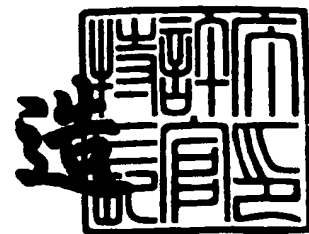
セイコーエプソン株式会社

CERTIFIED COPY OF
PRIORITY DOCUMENT

2001年 9月11日

特 許 庁 長 官
Commissioner,
Japan Patent Office

及 川 耕 造



出証番号 出証特2001-3083919

【書類名】 特許願

【整理番号】 J0080725

【提出日】 平成12年 9月20日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 3/00 515
G10L 3/00 535

【発明者】

【住所又は居所】 長野県諏訪市大和3丁目3番5号 セイコーエプソン株式会社内

【氏名】 宮沢 康永

【特許出願人】

【識別番号】 000002369

【氏名又は名称】 セイコーエプソン株式会社

【代表者】 安川 英昭

【代理人】

【識別番号】 100093388

【弁理士】

【氏名又は名称】 鈴木 喜三郎

【連絡先】 0 2 6 6 - 5 2 - 3 1 3 9

【選任した代理人】

【識別番号】 100095728

【弁理士】

【氏名又は名称】 上柳 雅誉

【選任した代理人】

【識別番号】 100107261

【弁理士】

【氏名又は名称】 須澤 修

【手数料の表示】

【予納台帳番号】 013044

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9711684

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 HMMの出力確率計算方法および音声認識装置

【特許請求の範囲】

【請求項1】 出力確率分布として無相関正規分布を用いたHMMの出力確率の計算方法において、ある時刻における入力ベクトルをY（Yはn次元の成分を有し、nは正の整数である）、状態iからjの場合のk次元（kは1からnのいずれかの自然数）における分散を $\sigma_{ij}(k)$ 、状態iからjの場合のk次元における平均値 $\mu_{ij}(k)$ としたとき、前記出力確率を求める計算式が、

【数1】

$$b_{ij}(Y) = \prod_{k=1}^n \left(\frac{1}{2\pi\sigma_{ij}(k)^2} \right)^{\frac{1}{2}} \cdot e^{-\sum_{k=1}^n \left[\frac{\{y_k - \mu_{ij}(k)\}^2}{2\sigma_{ij}(k)^2} \right]} \quad (1)$$

で与えられ、この式の両辺を対数をとって、

【数2】

$$\log_x b_{ij}(Y) = \underbrace{\log_x \left[\prod_{k=1}^n \left(\frac{1}{2\pi\sigma_{ij}(k)^2} \right)^{\frac{1}{2}} \right]}_A - \underbrace{\sum_{k=1}^n \left[\frac{\{y_k - \mu_{ij}(k)\}^2}{2\sigma_{ij}(k)^2} \right] \cdot \log_x e}_{B'} \quad (2)$$

で表した場合、(2)式の $\left[\{y_k - \mu_{ij}(k)\}^2 \right] / 2\sigma_{ij}(k)^2 \cdot \log_x e$ でなる計算項の計算量を少なくするHMMの出力確率計算方法であって、

ある時刻における入力ベクトルYの各次元ごとの値を代表値の集合として表すコードブックを作成し、前記ある時刻における入力ベクトルYの各次元（1～n

次元) ごとの値を前記コードブックに存在するあるコードで置換して、そのコードに基づいて、前記 $\{ \{ y_k - \mu_{ij}(k) \}^2 \} / 2 \sigma_{ij}(k)^2 \cdot \log_x e$ でなる計算項をテーブル参照で求め、

そのテーブルは、1 から n 次元のうち同じ組とすることのできる次元同志をまとめ、そのまとめられた組ごとに作成され、このそれぞれの組ごとのテーブルは、当該組に存在する各次元ごとに選ばれるそれぞれのコードに基づいて求められるそれぞれの出力値が、前記それぞれのコードの全ての組み合わせごとに合計の出力値として求められており、前記それぞれのコードの全ての組み合わせと、それによって得られる合計の出力値とが対応付けられてなり、

ある組に対応するそれぞれの次元に対して選ばれたそれぞれのコードに基づいて前記テーブルを参照し、それぞれの次元に対するコードの組み合わせに対応した合計の出力値を取得して、その合計の出力値に基づいて出力確率を計算することを特徴とする HMM の出力確率計算方法。

【請求項 2】 前記同じ組とすることのできる次元同志のまとまりは、それぞれの次元におけるコード数を次元によって異ならせることができる場合、コード数の同じ次元同志を同じ組としてまとめることを特徴とする請求項 1 記載の HMM の出力確率計算方法。

【請求項 3】 入力音声信号に対し特徴分析を行い、ある時刻ごとの入力ベクトルと複数次元からなる成分を出力する特徴分析部と、この特徴分析部からの各次元ごとの成分を受け取ってスカラ量子化を行い、所定のコードに置き換えるスカラ量子化部と、このスカラ量子化部によってコード化されたコードによって、予め作成されたテーブルを参照することで出力値を得て、その出力値を用いて出力確率を求めると共に、その出力確率を用いて音声認識に必要な演算を行う演算処理部と、この演算処理部で計算された結果に基づいて音声認識を行って認識結果を出力する音声認識処理部とを有する音声認識装置において、

前記テーブルは、1 から n 次元のうち同じ組とすることのできる次元同志をまとめ、そのまとめられた組ごとに作成され、このそれぞれの組ごとのテーブルは、当該組に存在する各次元ごとに選ばれるそれぞれのコードに基づいて求められるそれぞれの出力値が、前記それぞれのコードの全ての組み合わせごとに合計の

出力値として求められており、前記それぞれのコードの全ての組み合わせと、それによって得られる合計の出力値とが対応付けられてなり、

前記演算処理部は、ある組に対応するそれぞれの次元に対して選ばれたそれぞれのコードに基づいて前記テーブルを参照し、それぞれの次元に対するコードの組み合わせに対応した合計の出力値を取得して、その合計の出力値に基づいて出力確率を計算すると共に、当該計算された出力確率を用いて音声認識に必要な演算を行う、

ことを特徴とする音声認識装置。

【請求項 4】 前記同じ組とすることのできる次元同志のまとまりは、それぞれの次元におけるコード数を次元によって異ならせることができる場合、コード数の同じ次元同志を同じ組としてまとめることを特徴とする請求項 3 載の音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は音声認識において用いられる HMM（隠れマルコフモデル）の出力確率を少ない計算量で高速に求めることを可能とする HMM の出力確率計算方法および音声認識装置に関する。

【0002】

【従来の技術】

音声認識を行うための音素モデルとして HMM が広く使われている。この HMM は高い音声認識性能を得ることができる反面、計算量が多いのが難点である。特に、HMM の出力確率を求めるには多くの計算量を必要とする。ここで、ある時刻における入力ベクトル Y について、状態 i から状態 j へ遷移する時の出力確率を $b_{ij}(Y)$ で表し、無相関正規分布に従うとすれば、 $b_{ij}(Y)$ は前述の特許請求の範囲の項で示した (1) 式で表すことができる。

【0003】

このとき、入力ベクトル Y は、入力音声のある時刻（時刻 t_1 ，時刻 t_2 ， \cdots ）ごとに、たとえば 20 msec というような長さで分析して得られた n 次元（

n は正の整数)の成分(LPCケプストラムなど)で表される。たとえば、時刻 t_1, t_2, t_3, \dots における入力ベクトルを Y_1, Y_2, Y_3, \dots で表すとすれば、時刻 t_1 における入力ベクトル Y_1 は $(1y_1, 1y_2, \dots, 1y_n)$ 、時刻 t_2 における入力ベクトル Y_2 は $(2y_1, 2y_2, \dots, 2y_n)$ 、時刻 t_3 における入力ベクトル Y_3 は $(3y_1, 3y_2, \dots, 3y_n)$ というように表される。

【0004】

なお、上述の(1)式において、 k はある時刻における入力ベクトル Y の成分の次元数を表し、 1 から n のいずれかの値をとる。また、 $\sigma_{ij}(k)$ は状態 i から j の場合の k 次元における分散、 $\mu_{ij}(k)$ は状態 i から j の場合の k 次元における平均値を表す。

【0005】

【発明が解決しようとする課題】

上述した(1)式によって出力確率を求めることができるが、この(1)式をそのまま用いて計算すると、計算によって得られる値が小さくなりすぎてアンダーフローが起こる可能性がある。そこで、出力確率を求める際は、通常、対数化して求めることがなされている。上述の(1)式を x を底とする対数として表せば、前述の(2)式のように表すことができる

この(2)式において、計算項 A は、予め計算して求めておくことができるので、それを定数化して A と表し、さらに、計算項 B に存在する $\log_x e$ も定数として表すことができるので、それを Z で表せば、(2)式は、

【0006】

【数 3】

$$\log_x b_{ij}(Y) = A - \underbrace{\sum_{k=1}^n \left[\frac{\{y_k - \mu_{ij}(k)\}^2}{2\sigma_{ij}^2(k)} \cdot Z \right]}_{B} \quad (3)$$

と表すことができる。

【0007】

しかし、この(3)式においても計算量の多い計算項 B' 、つまり、 $\left[\frac{\{y_k - \mu_{ij}(k)\}^2}{2\sigma_{ij}^2(k)} \cdot Z \right]$ が存在する。特に、この項 B' は、ある時刻 t における入力ベクトル Y の各次元ごとに計算する必要がある。たとえば、ある時刻 t における入力ベクトル Y が10次元($n=10$)の成分で構成されているとすれば、10回の引き算、10回の掛け算、10回の割り算、10回の足し算を行い、さらに、定数 Z を掛け算する必要があり、これだけでもその計算量はきわめて多くなる。

【0008】

特に、小型、軽量、安価な製品においては、使用されるハードウェアに大きな制約があるため、上述したようなHMMを用いた音声認識を用いるには無理がある。

【0009】

そこで本発明は、HMM(隠れマルコフモデル)の出力確率を求めるための計算の簡略化を図り、出力確率を高速に求めることを可能とし、ハードウェアの能力に大きな制約のあるシステムにもHMMの適用を可能とすることを目的とする。

【 0 0 1 0 】

【課題を解決するための手段】

上述した目的を達成するために本発明のHMMの出力確率計算方法は、出力確率分布として無相関正規分布を用いたHMMの出力確率の計算方法において、ある時刻における入力ベクトルを Y （ Y は n 次元の成分を有し、 n は正の整数である）、状態 i から j の場合の k 次元（ k は1から n のいずれかの自然数）における分散を $\sigma_{ij}(k)$ 、状態 i から j の場合の k 次元における平均値 $\mu_{ij}(k)$ としたとき、前記出力確率を求める計算式が、（1）式で与えられ、この（1）式の両辺を対数を取り、（2）のように表した場合、（2）式の $\left[\{y_k - \mu_{ij}(k)\}^2 \right] / 2 \sigma_{ij}(k)^2 \cdot \log_x e$ でなる計算項の計算量を少なくするHMMの出力確率計算方法であって、ある時刻における入力ベクトル Y の各次元ごとの値を代表値の集合として表すコードブックを作成し、前記ある時刻における入力ベクトル Y の各次元（1～ n 次元）ごとの値を前記コードブックに存在するあるコードで置換して、そのコードに基づいて、前記 $\left[\{y_k - \mu_{ij}(k)\}^2 \right] / 2 \sigma_{ij}(k)^2 \cdot \log_x e$ でなる計算項をテーブル参照で求める。そして、そのテーブルは、1から n 次元のうち同じ組とすることのできる次元同志をまとめ、そのまとめられた組ごとに作成され、このそれぞれの組ごとのテーブルは、当該組に存在する各次元ごとに選ばれるそれぞれのコードに基づいて求められるそれぞれの出力値が、前記それぞれのコードの全ての組み合わせごとに合計の出力値として求められており、前記それぞれのコードの全ての組み合わせと、それによって得られる合計の出力値とが対応付けられてなり、ある組に対応するそれぞれの次元に対して選ばれたそれぞれのコードに基づいて前記テーブルを参照し、それぞれの次元に対するコードの組み合わせに対応した合計の出力値を取得して、その合計の出力値に基づいて出力確率を計算するようにしている。

【 0 0 1 1 】

このHMMの出力確率計算方法において、前記同じ組とすることのできる次元同志のまとまりは、それぞれの次元におけるコード数を次元によって異ならせることができる場合、コード数の同じ次元同志を同じ組としてまとめるようにしている。

【 0 0 1 2 】

また、本発明の音声認識装置は、入力音声信号に対し特徴分析を行い、ある時刻ごとの入力ベクトルと複数次元からなる成分を出力する特徴分析部と、この特徴分析部からの各次元ごとの成分を受け取ってスカラ量子化を行い、所定のコードに置き換えるスカラ量子化部と、このスカラ量子化部によってコード化されたコードによって、予め作成されたテーブルを参照することで出力値を得て、その出力値を用いて出力確率を求めると共に、その出力確率を用いて音声認識に必要な演算を行う演算処理部と、この演算処理部で計算された結果に基づいて音声認識を行って認識結果を出力する音声認識処理部とを有する音声認識装置において、前記テーブルは、1 から n 次元のうち同じ組とすることのできる次元同志をまとめ、そのまとめられた組ごとに作成され、このそれぞれの組ごとのテーブルは、当該組に存在する各次元ごとに選ばれるそれぞれのコードに基づいて求められるそれぞれの出力値が、前記それぞれのコードの全ての組み合わせごとに合計の出力値として求められており、前記それぞれのコードの全ての組み合わせと、それによって得られる合計の出力値とが対応付けられてなる。そして、前記演算処理部は、ある組に対応するそれぞれの次元に対して選ばれたそれぞれのコードに基づいて前記テーブルを参照し、それぞれの次元に対するコードの組み合わせに対応した合計の出力値を取得して、その合計の出力値に基づいて出力確率を計算すると共に、当該計算された出力確率を用いて音声認識に必要な演算を行う。

【 0 0 1 3 】

このように本発明は、HMMの出力確率を求める計算の一部、特に計算量の多い部分をテーブル参照で求められるようにしているので、出力確率を求めるための演算を簡略化することができ、出力確率を高速にかつ少ない処理量で求めることを可能としている。これにより、処理能力の低いCPUでも対応することができる。

【 0 0 1 4 】

また、本発明は、入力ベクトルの各次元を幾つかにまとめて、そのまとまりごとにテーブルを作成するようにしているので、テーブル数を少なくすることができ、それぞれのテーブルを参照することによって得られた値に対する後処理をよ

り簡略化することができる。

【0015】

また、次元同志のまとまりを作る際、各次元に対応するコードブック内のコード数の同じ次元同志を同じ組としてまとめるようにしている。これは、ある時刻の入力ベクトルがたとえば10次元の成分を持つと仮定すれば、1次元や2次元といった次元の数値が小さい次元では、コードブックのコード数をある程度多く密に持つ必要があるが、8次元、9次元、10次元というように数値が多くなるほどコードブックのコード数を疎とすることが可能となるというように、それぞれの次元におけるコード数を次元によって異ならせることができることを利用している。

【0016】

たとえば、1次元と2次元に対するコードブックのコード数は16個、3次元から6次元に対するコードブックのコード数はそれぞれ8個、7次元から10次元に対するコードブックのコード数はそれぞれ4個というように、それぞれの次元によってコード数を変えることができる場合、1次元と2次元を1つのまとまりとして、それに対応するテーブルを作成し、3次元から6次元を1つのまとまりとして、それに対応するテーブルを作成し、7次元から10次元を1つのまとまりとして、それに対応するテーブルを作成することができる。

【0017】

【発明の実施の形態】

以下、本発明の実施の形態について説明する。なお、この実施の形態で説明する内容は、本発明のHMMの出力確率計算方法および音声認識装置についての説明である。

【0018】

本発明は、前述したように、HMMの出力確率を求めるための計算の簡素化を図ろうとするものである。すなわち、本発明は簡単にいえば、上述の(3)式において、計算項B'をテーブル化して、そのテーブルを参照することで計算項B'を求めることができるようにして、HMMの出力確率の計算を簡素化しようとするものである。

【0019】

以下、詳細に説明する。なお、この実施の形態では、各時刻 t_1 , t_2 , t_3 , \dots ごとの入力ベクトル（これを Y_1 , Y_2 , Y_3 , \dots と表す）は、それぞれ 10 次元の成分で構成されているものとする。たとえば、時刻 t_1 における入力ベクトル Y_1 は $1y_1$, $1y_2$, \dots , $1y_n$ 、時刻 t_2 における入力ベクトル Y_2 は $2y_1$, $2y_2$, \dots , $2y_n$ 、時刻 t_3 における入力ベクトル Y_3 は $3y_1$, $3y_2$, \dots , $3y_n$ というような成分を有している。

【0020】

まず、1 から 10 次元の各次元ごとにそれぞれの次元ごとの値を代表値の集合として表すコードブックを持つ。これを図 1 に示す。図 1 は時刻 t_1 における入力ベクトル Y_1 のそれぞれの成分 $1y_1$, $1y_2$, \dots , $1y_{10}$ に対応するコードブックを示すもので、1 次元の成分 $1y_1$ に対しては、 $1C_1$, $1C_2$, \dots , $1C_m$ のコードを有するコードブック $1C$ 、2 次元においては、 $2C_1$, $2C_2$, \dots , $2C_m$ のコードブック $2C$ 、3 次元においては、 $3C_1$, $3C_2$, \dots , $3C_m$ のコードブック $3C$ 、10 次元においては、 $10C_1$, $10C_2$, \dots , $10C_m$ のコードブック $10C$ というように、それぞれの次元ごとのコードブック $1C$, $2C$, \dots , $10C$ を持つ。

【0021】

なお、ここで各コードブックのサイズを 64 とすれば、それぞれのコードブックにおける m は 64 となり、それぞれ 64 個のコードを有することになる。ただし、実際には、全ての次元が同じコード数を持つ必要はなく、それぞれの次元ごとにコード数を変えることができる。

【0022】

このように、各次元対応のコードブックを持つことによって、時刻 t_1 , t_2 , \dots , t_n のある時刻における入力ベクトル Y_1 , Y_2 , \dots , Y_n のそれぞれの 10 次元の成分は、それぞれの次元ごとのコードブック内に存在するいずれかのコードに置き換えることができる。

【0023】

たとえば、時刻 t_1 における入力ベクトル Y_1 における 1 次元目の値は、コー

ドブック 1 C の中から最も近い値を選ぶことにより、その選ばれた値に置き換えることができ、同じく入力ベクトル Y 1 における 2 次元目の値は、コードブック 2 C の中から最も近い値を選ぶことにより、その選ばれた値に置き換えることができる。

【 0 0 2 4 】

同様にして、時刻 t 2 の入力ベクトル Y 2 における 1 次元目 2 y 1 の値も、コードブック 1 C の中から最も近い値を選ぶことにより、その選ばれた値に置き換えることができるというように、各時刻における入力ベクトルのそれぞれの次元の成分は、それぞれ対応するコードブック内の m 個のコードのいずれかに置き換えることができる。

【 0 0 2 5 】

そして、このようにそれぞれのコードブックの中のいずれかのコードに置き換えられると、(3) 式は以下のように表すことができる。

【 0 0 2 6 】

【数 4】

$$\log_x b_{ij}(Y) = A - \underbrace{\sum_{k=1}^n \left[\frac{\{kC_c - \mu_{ij}(k)\}^2}{2\sigma_{ij}^2(k)} \cdot Z \right]}_E \quad (4)$$

この (4) 式において、k C c は置き換えられたコードを示すもので、k は次元を表し、1 から n までの次元があるとすれば、k は 1 から n のいずれかの値を取り、c はコード番号を表しコードブックサイズが m 個であれば 1 から m のいずれかの値をとる。

【0027】

この(4)式の計算項Dは、各次元ごとの値として予め求めておくことができ、それによって出力テーブルを作成する。すなわち、(4)式における計算項Dは、「 $\sigma_{ij}(k)$ 」、「 $\mu_{ij}(k)$ 」、「 kCc 」、「Z」からなり、「 $\sigma_{ij}(k)$ 」は前述したように、状態iからjの場合のk次元における分散、 $\mu_{ij}(k)$ は状態iからjの場合のk次元における平均値を表すもので、予め計算して求めておくことが可能であり、Zは定数化されている。また、 kCc はそれぞれの次元ごとのコードブック1C, 2C, ..., nCに存在するコードであるのでそれを用いることができる。

【0028】

これら各次元ごとの「 $\sigma_{ij}(k)$ 」、「 $\mu_{ij}(k)$ 」、定数化されている「Z」を用い、かつ、その次元に対するコードブックに存在するm個のコードを用いて計算項Dを計算しておくことができる。たとえば、コードブック1C内の各コード1C1~1Cm ($m=64$)をそれぞれ用い、(4)式の計算項Dを予め計算しておくことによって、各コード1C1から1Cmのそれぞれに対する計算項Dの計算結果(これをここでは出力値と呼ぶ)を得ることができ、それをテーブル化する(これを1次元のテーブルT1という)。

【0029】

また、コードブック2C内の各コード2C1~2Cm ($m=64$)をそれぞれ用い、(4)式の計算項Dを予め計算しておくことによって、各コード2C1から2Cmのそれぞれに対する計算項Dの出力値を得ることができ、それをテーブル化する(これを2次元のテーブルT2という)。

【0030】

このようにして、1~nの各次元に対応したn個の出力テーブル(T1~Tn)が作成され、これら各テーブルT1~Tnは、1~nの各次元ごとにm個のコードに対応したm個の出力値を持つことになる。

【0031】

したがって、ある時刻における入力ベクトルYが得られて、その入力ベクトルYの各次元ごとの成分が得られれば、各次元ごとのスカラ量子化されたコードが

得られ、そのコードに基づいてその次元に対応するテーブルを参照すれば（４）式の計算項Dが求められることになる。

【0032】

たとえば、ある時刻 t_1 における入力ベクトル Y_1 の1次元目の成分 $1y_1$ に対し、コード1C3が選ばれたとすると、そのコード1C3を用いて1次元のテーブルT1を参照することにより、計算項Dにおける1次元目の出力値が求められることになり、同様に、ある時刻 t_1 における入力ベクトル Y_1 の2次元目の成分 $1y_2$ に対し、コード2C16が選ばれたとすると、そのコード2C16を用いて2次元のテーブルT2を参照することにより、計算項Dにおける2次元目の出力値が求められることになる。

【0033】

これを図2により説明する。同図（a）に示すように、ある時刻 t_1 における入力ベクトル Y_1 の1次元目の成分を $1y_1$ とし、この $1y_1$ によってコードブック1Cからコード1C3が選ばれたとする。そして、このコード1C3によって、1次元の出力テーブルT1を参照することで、出力値103を得る。この出力値103は、（４）式の計算項Dから予め求められているものである。

【0034】

同様にして、同図（b）に示すように、同じ入力ベクトル Y_1 の2次元目の成分を $1y_2$ とし、この $1y_2$ によってコードブック2Cからコード1C16が選ばれたとする。そして、このコード1C16によって、2次元のテーブルT2を参照することで、出力値2016を得る。この出力値2016は、（４）式の計算項Dから予め求められているものである。

【0035】

このようにして n 次元までそれを繰り返すことで、それぞれの次元ごとの出力値が得られる。そして、それぞれの次元ごとの出力値（ここでは、説明を簡単にするために、各次元ごとの出力値を $O_1, O_2, O_3, \dots, O_n$ で表す）を足し算（ $O_1 + O_2 + \dots + O_n$ ）する。この $O_1 + O_2 + \dots + O_n$ は（４）式の計算項Eの計算結果となるので、その足し算した結果を用いて（４）式全体を計算すれば、ある時刻の入力ベクトル Y の出力確率 b_{ij} （ Y

）が求められる。

【0036】

以上のような処理を行うことにより、(4)式の項Eを得るのに必要な計算は、1つの次元について1回のコードブックを参照する処理と、その結果に基づいてテーブルT1～Tnのいずれかのテーブルを参照する処理を行い、得られた出力値を足し算すればよく、たとえば10次元とした場合、それぞれの次元ごとの10回のコードブック参照と、10回のテーブル参照と、それによって得られた10個の出力値を足し算するだけの処理ですむ。

【0037】

以上は本発明を説明するための基本的な処理を説明するものであって、各次元ごとに作成されたテーブルT1～Tnを参照することにより(4)式の計算項Dを得て、それぞれに次元ごとの出力値を足し算することによって(4)式の計算項Eを得る。このような計算手法はすでに知られた方法であるが、本発明は、この考え方をさらに発展させたものであり、より効率よく少ない計算量で出力確率を求めることができるようにする。以下、本発明について説明する。

【0038】

本発明は、前述した出力テーブルを個々の次元ごとに持つのではなく、1からnの次元のうち同じ組とすることのできる次元同志をまとめ、そのまとめられた組ごとにテーブルをもつことに特徴がある。

【0039】

図3は本発明の出力確率計算手順の概略を示すフローチャートであり、入力ベクトルの1からnの次元のうち同じ組とすることのできる次元同志を幾つかにまとめて(ステップs1)、そのまとまりごとにテーブルを作成しておく(ステップs2)。そして、出力確率を計算する際は、入力ベクトルの1次元からn次元までの成分に対するコードを、対応するコードブックから順次得て(ステップs3)、それぞれのコードを基に、対応するテーブルを参照してテーブルごとの出力値を得る(ステップs4)。そして、そのテーブルごとの出力値を出力確率を求めるための計算式(前述の(4)式)に代入して出力確率を求める(ステップs5)。以下、詳細に説明する。

【 0 0 4 0 】

各時刻ごとの入力ベクトルのそれぞれの次元における成分をスカラ量子化したコードブックを生成する際、前述したように、それぞれの次元ごとにコード数を変えることができる。たとえば、各時刻の入力ベクトルが10次元の成分を持つと仮定すれば、1次元や2次元といった次元の数値が小さい次元では、コードブックのコード数を密に持ち、8次元、9次元、10次元というように次元の数値の大きな次元ではコードブックのコード数を疎とするようにしてもよい。

【 0 0 4 1 】

ここで、入力ベクトルが10次元のLPCケプストラム係数である場合、1つの例として、1次元に対するコードブックのコード数は16個、2次元に対しても16個、3次元に対するコードブックのコード数は8個、4次元に対しても8個、5次元、6次元に対してもそれぞれ8個、7次元に対するコードブックのコード数は4個、8次元に対しても4個、9次元、10次元に対してもそれぞれ4個というように、それぞれの次元によってコード数を変えることができる。

【 0 0 4 2 】

したがって、それぞれのまとまりを生成する際、この場合、1次元と2次元がともに16個のコード数、3次元から6次元がそれぞれ8個のコード数、7次元から10次元がそれぞれ4個のコード数であるから、同じコード数ごとに次元をまとめると都合がよい。なお、説明の都合上、1次元と2次元のまとまりを第1の組V1、3次元～6次元のまとまりを第2の組V2、7次元～10次元のまとまりを第3の組V3と呼ぶことにする。

【 0 0 4 3 】

そして、これら第1～第3の組V1、V2、V3ごとにテーブル（第1のテーブルT10、第2のテーブルT20、第3のテーブルT30）を作成する。

【 0 0 4 4 】

第1のテーブルT10は、ある時刻における入力ベクトルYにおける1次元の成分 y_1 に対してコードブック1Cの中から選ばれたあるコード（これを1C_mとする）に対する出力値（これを1O_mとする）と、同じ時刻の入力ベクトルYの2次元の成分 y_2 に対してコードブック2Cの中から選ばれたあるコード（こ

れを $2C_m$ とする) に対する出力値 (これを $2O_m$ とする) とを足し算した結果 ($1O_m + 2O_m$) が、コードブック $1C$ と $2C$ のそれぞれに存在する全てのコードの組み合わせに対応して格納されているものである。

【0045】

つまり、コードブック $1C$ に存在する全てのコードに対して (4) 式の計算項 D を求めておくとともに、コードブック $2C$ に存在する全てのコードに対して (4) 式の計算項 D を求めておく。そして、コードブック $C1$ と $C2$ に存在するコードの全ての組み合わせに対するそれぞれの出力値 $1O_m + 2O_m$ を予め計算しておく。これを図4によって具体的に説明する。

【0046】

たとえば、時刻 t_1 における入力ベクトル Y_1 の1次元の成分 $1y_1$ に対してコードブック $1C$ からコード $1C3$ が選ばれ、同時刻における入力ベクトル Y_1 の2次元の成分 $1y_2$ に対してコードブック $2C$ からコード $2C16$ が選ばれたとする。

【0047】

このとき、コード $1C3$ に対する (4) 式の計算項 D はすでに求められていてその出力値を $1O3$ 、コード $2C16$ に対する (4) 式の計算項 D もすでに求められていてその出力値を $2O16$ とする。そして、これらの出力値の合計値「 $1O3 + 2O16$ 」もすでに計算されていて ($1O3 + 2O16 = 01$ とする)、その合計値は第1のテーブル $T10$ に格納されている。したがって、1次元の成分 $1y_1$ に対して選ばれたコード $1C3$ と、2次元の成分 $1y_2$ に対して選ばれたコード $2C16$ とによって第1のテーブル $T10$ を参照すると、この場合、「 $1O3 + 2O16$ 」、つまり、出力値として「 01 」が読み出される。

【0048】

これは他の組に対する出力テーブルも同じで、それぞれの組に属する次元ごとにその次元のコードブックに存在する全てのコードに対し、(4) 式の項 D がすでに求められていて、それぞれの組に属する次元ごとの全てのコードの組み合わせに対する合計の出力値が算出されている。

【0049】

たとえば、図5に示すように、時刻 t_1 における入力ベクトル Y_1 の3次元の成分 $1y_3$ に対してコードブック3Cからコード3C5が選ばれ、同時刻における入力ベクトル Y_1 の4次元の成分 $1y_4$ に対してコードブック4Cからコード4C6が選ばれたとする。また、同時刻における入力ベクトル Y_1 の5次元の成分 $1y_5$ に対してコードブック5Cからコード5C7が選ばれ、さらに、同時刻における入力ベクトル Y_1 の6次元の成分 $1y_6$ に対してコードブック6Cからコード6C8が選ばれたとする。

【0050】

このとき、コード3C5に対しては(4)式の計算項Dにより、その出力値として305が求められており、コード4C6に対しては(4)式の計算項Dにより、その出力値として406が求められており、コード5C7に対しては(4)式の計算項Dにより、その出力値として507が求められており、コード6C8に対しては(4)式の計算項Dにより、その出力値として608が求められているとする。

【0051】

そして、 $305 + 406 + 507 + 608$ もすでに計算されていて($305 + 406 + 507 + 608 = 02$ とする)、その合計値は第2のテーブルT20に格納されている。したがって、3次元の成分に対するコード3C5、4次元の成分に対するコード4C6、5次元の成分に対するコード5C7、6次元の成分に対するコード6C8によって、第2のテーブルT20を参照すると、この場合、「 $305 + 406 + 507 + 608$ 」、つまり、出力値として「02」が読み出される。

【0052】

さらに、図6に示すように、時刻 t_1 における入力ベクトル Y_1 の7次元の成分 $1y_7$ に対してコードブック7Cからコード7C1が選ばれ、同時刻における入力ベクトル Y_1 の8次元の成分 $1y_8$ に対してコードブック8Cからコード8C2が選ばれたとする。また、同時刻における入力ベクトル Y_1 の9次元の成分 $1y_9$ に対してコードブック9Cからコード9C3が選ばれ、さらに、同時刻における入力ベクトル Y_1 の10次元の成分 $1y_{10}$ に対してコードブック10C

からコード10C4が選ばれたとする。

【0053】

このとき、コード7C1に対しては(4)式の計算項Dにより、その出力値として701が求められており、コード8C2に対しては(4)式の計算項Dにより、その出力値として802が求められており、コード9C3に対しては(4)式の計算項Dにより、その出力値として903が求められており、コード10C4に対しては(4)式の計算項Dにより、その出力値として1004が求められているとする。

【0054】

そして、 $701 + 802 + 903 + 1004$ もすでに計算されていて($701 + 802 + 903 + 1004 = 03$ とする)、その合計値は第3のテーブルT30に格納されている。したがって、7次元の成分に対するコード7C1、8次元の成分に対するコード8C2、9次元の成分に対するコード9C3、10次元の成分に対するコード10C4によって、第3のテーブルT30を参照すると、この場合、「 $701 + 802 + 903 + 1004$ 」、つまり、出力値として「03」が読み出される。

【0055】

以上のように、1から10の次元を3つのまとまり(第1の組V1、第2の組V2、第3の組V3)に分けて、これら第1の組V1、第2の組V2、第3の組V3ごとにテーブル(第1のテーブルT10、第2のテーブルT20、第3のテーブルT30)を持つようにしている。したがって、第1、第2、第3の組V1、V2、V3に属する次元のコードブックから選ばれたそれぞれの次元ごとのコードが対応するテーブルに入力されると、入力されたコード群の組み合わせに対応する合計の出力値が得られる。

【0056】

これによって、たとえば、第1の組V1に対応するテーブルT10からの出力値が01、第2の組V2に対応するテーブルT20からの出力値が02、第3の組V3に対応するテーブルT30からの出力値が03であるとすれば、これら3つの出力値01、02、03を合計することで、(4)式の計算項Eの計算が済

むことになる。すなわち、(4) 式の計算項 E を得るための処理としては、10 次元とした場合、10 回のコードブック参照と、3 回のテーブル参照と、3 回の足し算だけで済むことになり、大幅に演算量を減らすことができる。

【0057】

そして、(4) 式の計算項 E が求められれば、それを用いて (4) 式全体を計算することによって、時刻 t における入力ベクトルの状態 i から状態 j の出力確率を求めることができる。

【0058】

このように、本発明によれば、出力確率を求める際、(4) 式の計算項 E を求めるに必要な処理は、各次元ごとのコードブック参照と、それによって得られたコードを用いて次元ごとにまとめられた幾つかの（上述の例では 3 つ）テーブルを参照して、それによって得られた値を足し算するだけの簡単処理で済むので、演算量を大幅に削減することができ、処理能力の低い CPU であっても十分対応できるものとなる。

【0059】

ところで、これまでの説明は、ある 1 つの正規分布に対しての説明であった。しかし、本発明は複数の正規分布を用いる混合連続分布型 HMM にも対応することができる。以下、これについて説明する。

【0060】

本発明では、全ての音素に共通の m 個のコード化された正規分布（代表的な m 個の正規分布がコード化されてコードブックとされているもので、それぞれのコードに対応する正規分布の分散と平均値が予め求められている）を予め用意し、ある音素について状態 i, j において時刻 t における入力ベクトル Y_t のある次元 y_k における出力確率を求めることができるようにする。

【0061】

この m 個（たとえば 256 個）のコード化された正規分布は、全ての音素に対して共通に使用可能であり、ある音素について状態 i, j ごとに、必要に応じて、コードブックから複数の正規分布を選択して出力確率を求めることができるようにしたものである。 m 番目の正規分布を用いた場合の出力確率を $b_m(Y)$ と

すれば、 $b_m(Y)$ は以下の式で表すことができる。

【0062】

【数5】

$$W_{ijm} \cdot b_m(Y) = W_{ijm} \cdot \prod_{k=1}^n \left(\frac{1}{2\pi\sigma_m^2(k)} \right)^{\frac{1}{2}} \cdot e^{-\sum_{k=1}^n \left[\frac{\{y_k - \mu_m(k)\}^2}{2\sigma_m^2(k)} \right]} \quad (5)$$

なお、(5)式において、 W_{ijm} は、それぞれの音素における状態 i, j における m 番目の正規分布に対する重みである。

【0063】

この(5)式は、前述したと同様に、対数をとって、

【0064】

【数6】

$$\begin{aligned} & \log_x W_{ijm} \cdot b_m \\ &= \underbrace{\log_x W_{ijm} + \log_x \left[\prod_{k=1}^n \left(\frac{1}{2\pi\sigma_m^2(k)} \right)^{\frac{1}{2}} \right]}_A - \underbrace{\sum_{k=1}^n \left[\frac{\{y_k - \mu_m(k)\}^2}{2\sigma_m^2(k)} \right] \cdot \log_x e}_{B'} \quad (6) \end{aligned}$$

と表すことができる。そして、この(6)式において、計算項Aは、前述したように、予め計算して求めておくことができるので、それを定数化してAと表し、さらに、計算項Bにおける $\log_x e$ も前述同様、定数として表すことができるので

、それをZで表せば、(6)式は、

【0065】

【数7】

$$\log_x W_{ijm} \cdot b_m = \log_x W_{ijm} + A - \underbrace{\sum_{k=1}^n \left[\frac{\{y_k - \mu_m(k)\}^2}{2\sigma_m^2(k)} \cdot Z \right]}_{B'} \quad (7)$$

B

と表すことができる。

【0066】

そしてさらに、前述同様、(7)式の計算項B'におけるy_kをスカラ量子化してコード化し、それをkC_cで表すと、(7)式は、

【0067】

【数8】

$$\log_x W_{ijm} \cdot b_m = \log_x W_{ijm} + A - \underbrace{\sum_{k=1}^n \left[\frac{\{kC_c - \mu_m(k)\}^2}{2\sigma_m^2(k)} \cdot Z \right]}_{D} \quad (8)$$

E

のように表すことができる。なお、(7)式、(8)式において、 $\log xW_{ijm}$ も予め計算して求めておくことができるので定数として扱うことができる。

【0068】

そして、この(8)式における計算項Dをテーブル化してそのテーブルを参照することで求めることができるようにする。これについてはすでに詳細に説明したのでここではその説明は省略する。

【0069】

このように本発明は、全ての音素に共通のコード化されたm個の正規分布を予め用意することで、ある音素について状態i jにおいて時刻における入力ベクトルのある次元における出力確率を(8)式によって簡単に求めることができる。

【0070】

図7は以上説明したHMMの出力確率計算方法が適用された音声認識装置の概略的な構成を示すもので、音声入力部としてのマイクロホン11、入力された音声信号を増幅したりA/D変換する音声信号入力処理部12、音声信号入力処理部12で処理された音声信号に対し特徴分析を行い時刻tごとの入力ベクトルとその成分(たとえば10次元のLPCケプストラム)を出力する処理などを行う特徴分析部13、この特徴分析部13からの各次元ごとの成分を受け取ってコードブック記憶部14に記憶されている次元ごとのコードブック(たとえば図1に示すコードブック1C, 2C, ..., 10C)を参照してスカラ量子化を行うスカラ量子化部15、HMM演算に必要な各種パラメータ(状態i jの場合の次元kにおける σ_{ij} や μ_{ij} など)や前述したテーブル(第1、第2、第3のテーブルT10, T20, T30)など各種データを記憶する演算データ記憶部16、この演算データ記憶部16に記憶されている各種データと上述のスカラ量子化部15によってスカラ量子化された結果(各次元の成分に対応するコードブックから選ばれたコード)に基づいてHMM演算を行う演算処理部17、この演算処理部17で計算された結果を用い、かつ、単語や言語テーブル18を用いて音声認識を行って認識結果を出力する音声認識処理部19などを有した構成となっている。

【0071】

なお、演算処理部 1 7 が行う処理の 1 つである出力確率を求める処理は、前述したように、スカラ量子化部 1 5 によって選ばれたコードに基づいて、テーブル（前述の例では第 1、第 2、第 3 のテーブル T 1 0、T 2 0、T 3 0）を参照して各テーブルごとの出力値を求め、それによって出力確率を計算して、この出力確率を用いて最終的な出力尤度を得る。これをある音素において各時刻ごとに行う。そして、この演算処理部 1 7 からの出力を音声認識処理部が受け取って入力音声に対する音声認識処理を行う。

【 0 0 7 2 】

このように本発明の音声認識装置は、図 1 から図 6 によって説明した HMM の出力確率計算方法を用いて出力確率を求めるようにしている。これによって、出力確率を求めるための演算が大幅に簡略化され、高速に出力確率を求めることができ、処理能力の低い CPU で十分対応できるものなる。これによって、低価格が要求される安価なシステムにおいても HMM を用いた音声認識を適用することが可能となる。

【 0 0 7 3 】

本発明は以上説明した実施の形態に限定されるものではなく、本発明の要旨を逸脱しない範囲で種々変形実施可能となるものである。たとえば、前述の実施の形態では、音声認識について説明したが、画像認識を行う際のパターン認識などにも適用することができる。また、以上説明した本発明の HMM の出力確率計算方法を実現するための処理手順が記述された処理プログラムを作成し、その処理プログラムをフロッピーディスク、光ディスク、ハードディスクなどの記録媒体に記録させておくことができ、本発明はその記録媒体をも含むものである。また、ネットワークから当該処理プログラムを得るようにしてもよい。

【 0 0 7 4 】

【発明の効果】

以上説明したように本発明によれば、HMM の出力確率を求める計算の一部、特に計算量の多い部分をテーブル参照で求められるようにしているので、出力確率を求めるための演算を簡略化することができ、出力確率を高速にかつ少ない処理量で求めることを可能としている。これにより、処理能力の低い CPU でも

対応することができる。また、本発明は、入力ベクトルの各次元を幾つかにまとめて、そのまとまりごとにテーブルを作成するようにしているので、テーブル数を少なくすることができ、それぞれのテーブルを参照することによって得られた値に対する後処理をより簡略化することができる。

【0075】

このように本発明は、HMMの出力確率を求める演算が大幅に簡略化され、高速に出力確率を求めることができ、処理能力の低いCPUで十分対応できるものなる。

【図面の簡単な説明】

【図1】

本発明の実施の形態を説明する図であり、時刻 t_1 における入力ベクトル Y_1 の各次元の成分をコード化する処理を説明する図である。

【図2】

図1に示されるコードブックによってコード化されたコードを用いてテーブル参照することにより、実施の形態で説明した(4)式の計算項Dを求める処理を説明する図である。

【図3】

本発明の実施の形態であるHMMの出力確率を求めるための概略的な処理手順を説明するフローチャートである。

【図4】

本発明の実施の形態で説明した(4)式の計算項Dをテーブル参照により求める処理(1次元と2次元に対する処理)を説明する図である。

【図5】

本発明の実施の形態で説明した(4)式の項Dをテーブル参照により求める処理(3次元から6次元に対する処理)を説明する図である。

【図6】

本発明の実施の形態で説明した(4)式の項Dをテーブル参照により求める処理(7次元と10次元に対する処理)を説明する図である。

【図7】

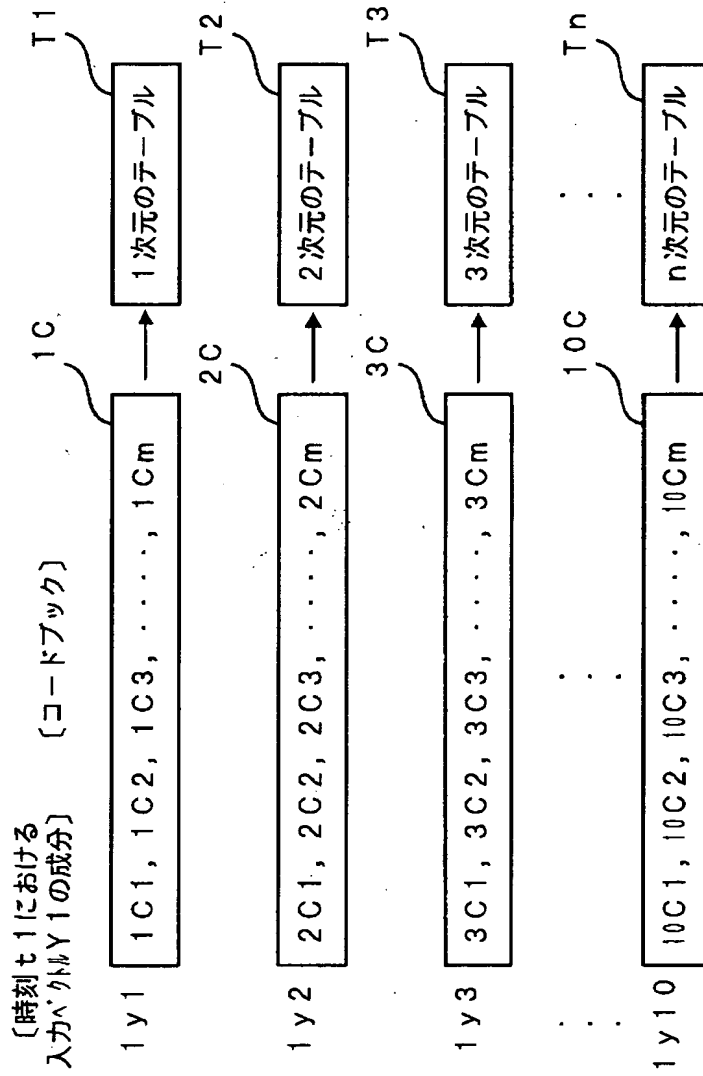
本発明の音声認識装置の実施の形態を概略的に示す構成図である。

【符号の説明】

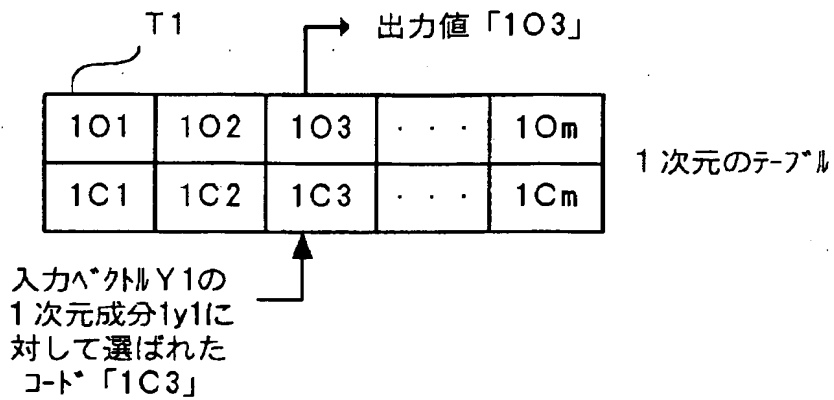
- 11 マイクロホン
- 12 音声信号入力処理部
- 13 特徴分析部
- 14 コードブック記憶部
- 15 スカラ量子化部
- 16 演算データ記憶部
- 17 演算処理部
- 18 音声認識処理部
- 19 単語・言語テーブル
- 1C, 2C, ..., 10C 1次元から10次元に対応するコードブック
- T10, T20, T30 第1、第2、第3のテーブル

【書類名】 図面

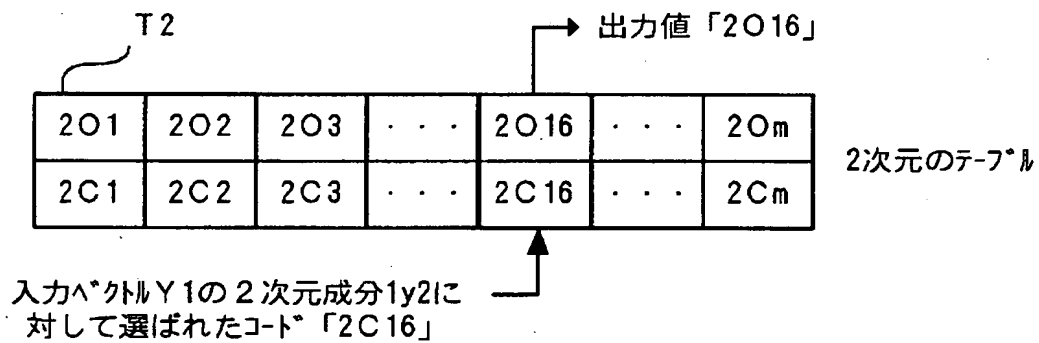
【図 1】



【図 2】

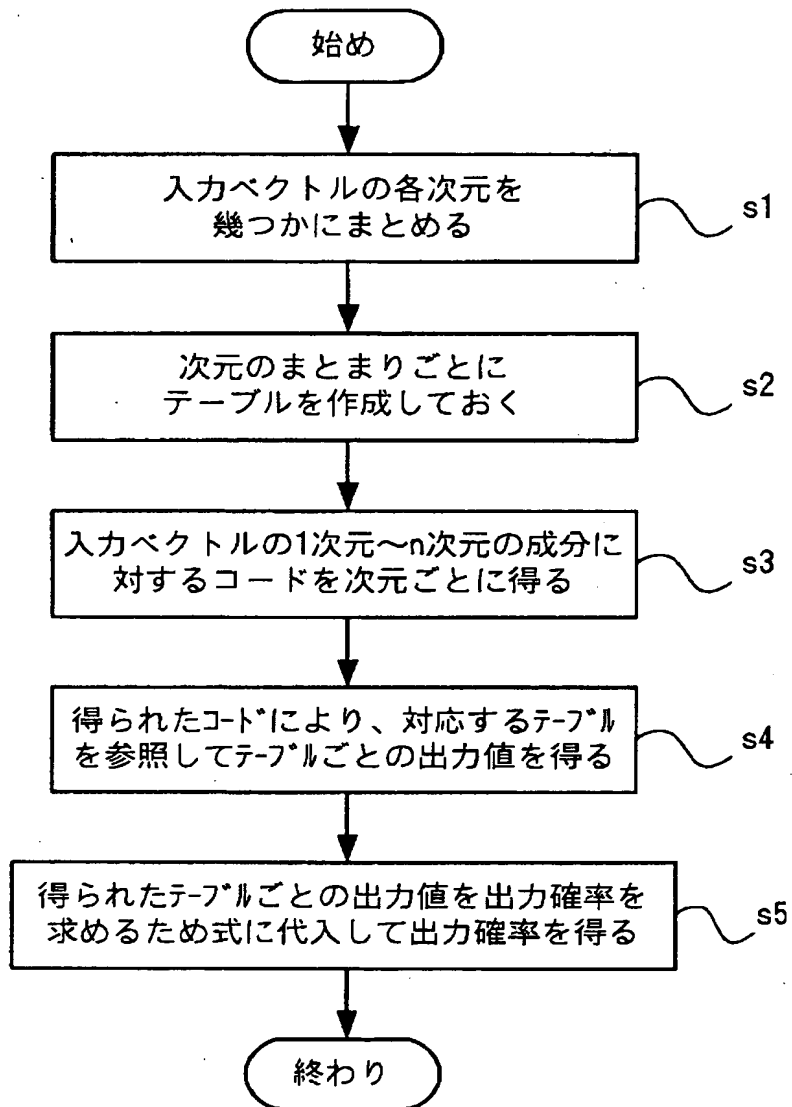


(a)

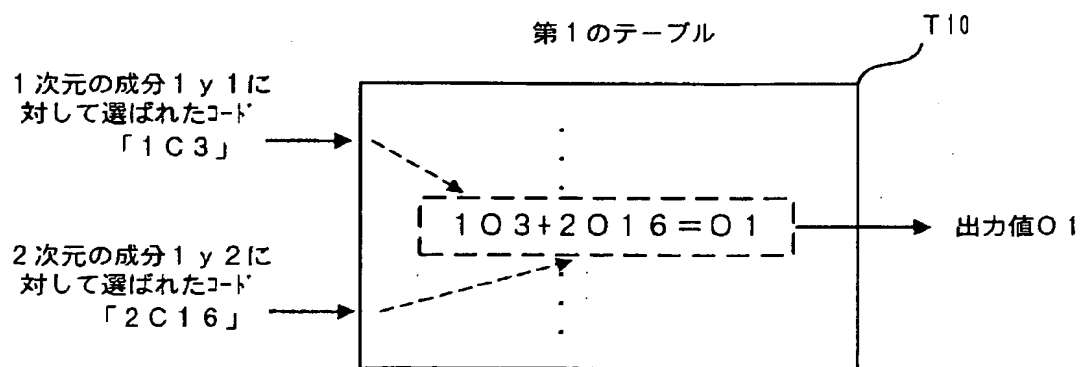


(b)

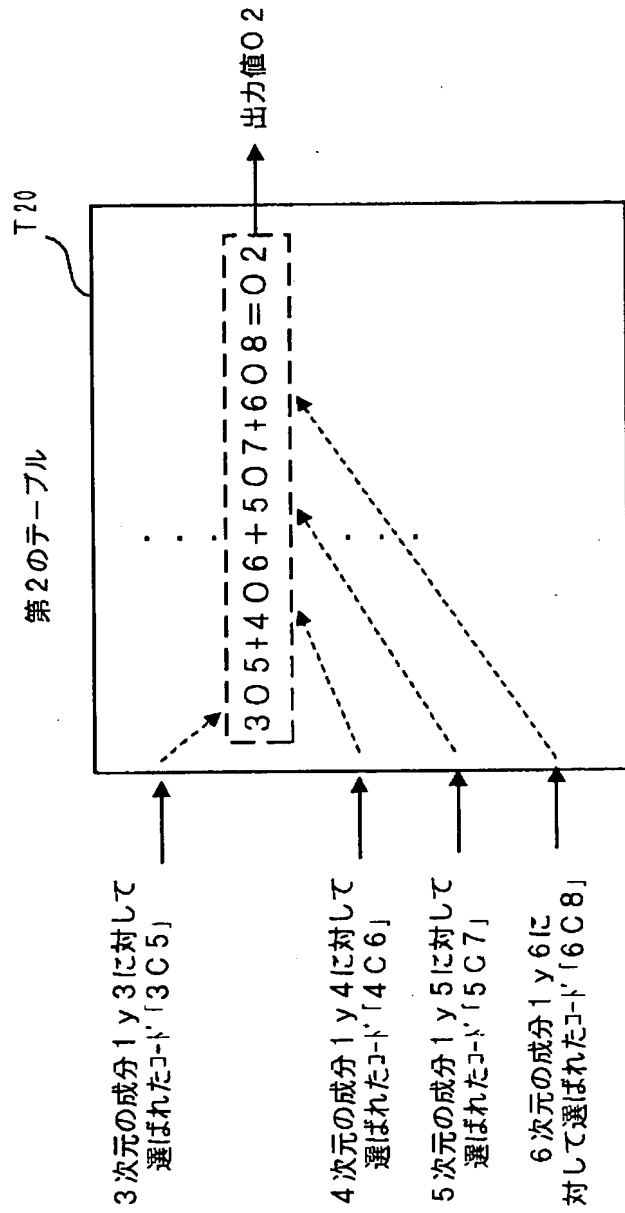
【図3】



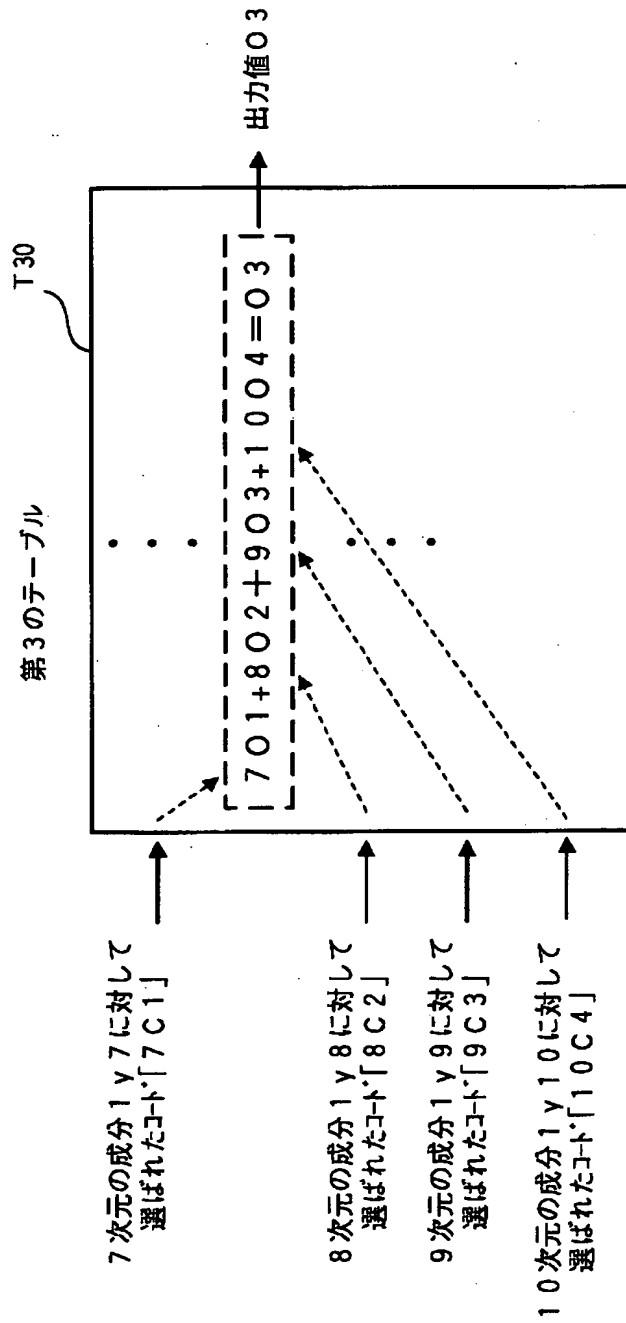
【図 4】



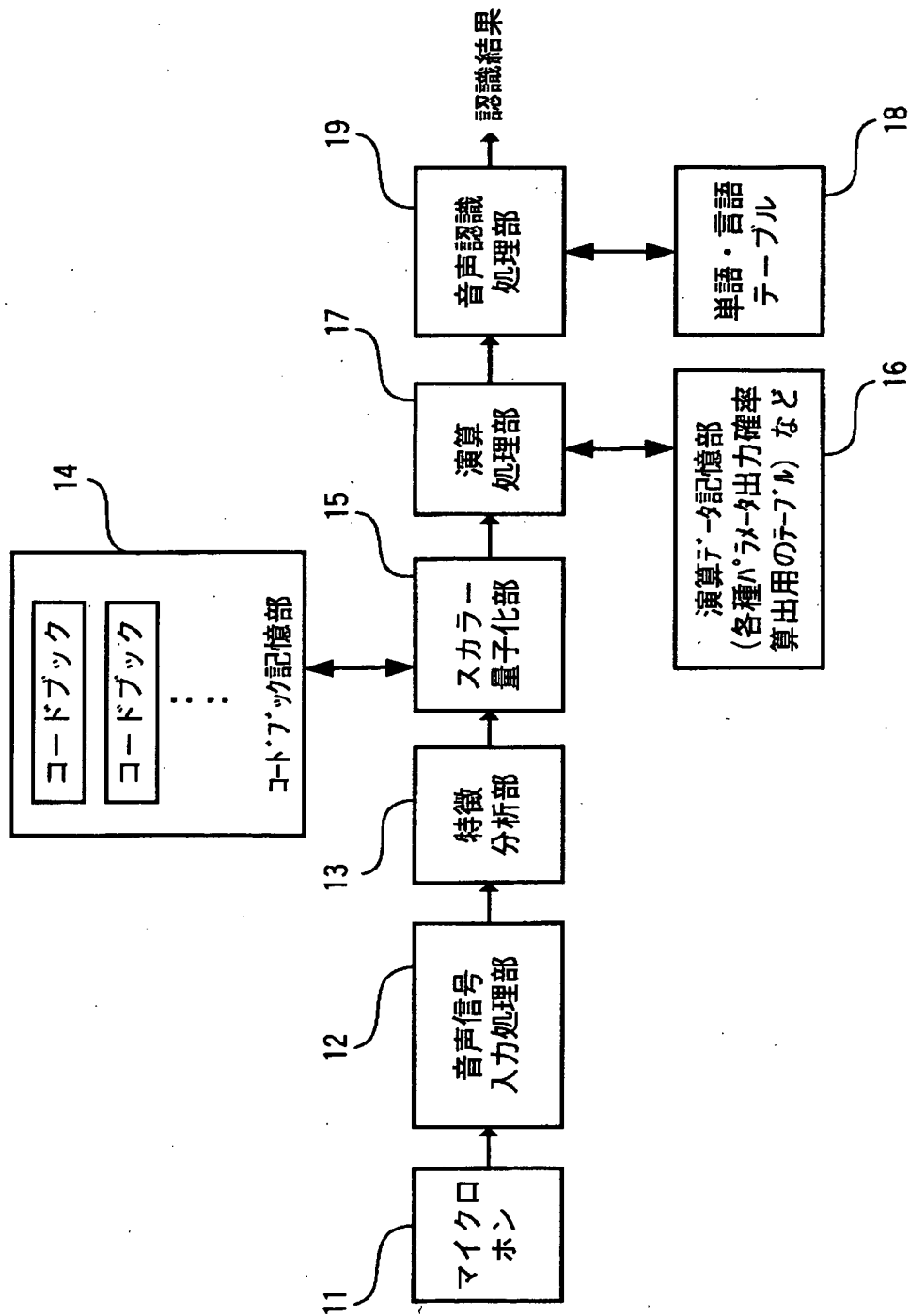
【図 5】



【図 6】



【図 7】



【書類名】 要約書

【要約】

【課題】 HMMの出力確率を求めるための演算を簡略化することで処理能力の低いCPUでも対応することができるようにする。

【解決手段】 入力ベクトルの各次元を幾つかにまとめて（ステップs 1）、そのまとまりごとにテーブルを作成しておく（ステップs 2）。そして、出力確率を計算する際は、入力ベクトルの1次元からn次元までの値に対するコードを順次得て（ステップs 3）、それぞれのコードによって、対応するテーブルを参照してテーブルごとの出力値を得る（ステップs 4）。そして、そのテーブルごとの出力値を出力確率を求めるための計算式に代入して出力確率を求める（ステップs 5）。

【選択図】 図 3

出 願 人 履 歴 情 報

識別番号 [000002369]

1. 変更年月日	1990年 8月20日
[変更理由]	新規登録
住 所	東京都新宿区西新宿2丁目4番1号
氏 名	セイコーエプソン株式会社